arXiv:physics/9908021 v1 10 Aug 1999

Microarrays for Amino Acid Analysis and Protein Sequencing

Edward Shipwash

NanoNano Technology, 1639 Addison St., Berkeley CA 94703

Summary

A method is described where the aminoacyl-tRNA synthetase system is used to create very small devices for quantitative analysis of the amino acids that occur in proteins. The basis of the method is that each of the 20 synthetases and/or a tRNA specific for a different amino acid is separated spatially (e.g. in tiny chambers or on a surface). The reactions catalyzed by all 20 synthetases are monitored in a spatially resolved manner. Each separately positioned synthetase or tRNA will signal its cognate amino acid. The synthetase reactions can be monitored using continuous spectroscopic assays. Alternatively, since elongation factor Tu;GTP (EF-Tu;GTP) specifically binds all AA-tRNAs, the aminoacylation reactions catalyzed by the synthetases can be monitored using ligand assays. Microarrays for amino acid analysis are suggested. Additionally, it is possible that amino acid analysis arrays can be integrated with aminopeptidase or carboxypeptidase digestions to produce miniaturized enzymatic sequenators capable of generating either N- or C-terminal sequence data at femtomole-attomole levels. The possibility of parallel processing of many samples in an automated manner is discussed.

Introduction

Microfabrication technology and position-sensitive detectors have revolutionized genome analysis. Microarrays detect gene expression levels in parallel by measuring the hybridization of mRNA to many thousands of genes immobilized at high spatial resolution on a glass surface (1). Highly resolved detection is generally achieved by laser induced fluorescence of a labeled probe. Results from massively parallel and quantitative gene expression measurements analyzing up to 40,000 genes at a time and whole-genome variant detection methods show the power and accuracy of combining biorecognition phenomenon with miniaturized array based methods (2). Capillary array electrophoresis where many capillaries are run and detected in parallel has recently been developed for rapid DNA sequencing (3). On-line microfluidic systems that transport liquid solutions in channels of micron dimensions have been used for high-throughput DNA genotyping (4), polymerase chain reactions, and DNA sequencing reactions(5).

Recent advances in rapid gene analysis have not, however, been paralleled for proteins. With the realization that physiology is a manifestation of protein chemistry, proteome projects, that aim to characterize all proteins expressed by a genome or tissue, have recently emerged(6). The identification of proteins separated by two-dimensional (2D) gel electrophoresis is pivotal to proteome projects. 2D gel electrophoresis is a process where thousands of proteins can be purified to homogeneity in a single experiment(7). Presently no method exist to identify these proteins rapidly and inexpensively. Often a protein can be identified on the basis of its amino acid composition (8) or by N- or C-terminal 'sequence tags' only three or four amino acids long (9), but current amino acid analyzers and protein sequenators are too slow, expensive, and insensitive for proteome projects. C-terminal sequence tags are more specific than N-terminal

tags the same length, but no reliable, sensitive method for C-terminal protein sequencing is currently available. Hence, new rapid methods for amino acid analysis and end-group protein sequencing are needed.

Here, I describe a new method for amino acid analysis based on biomolecular recognition phenomenon. The key recognition event collating the nucleic acid sequences of genes and amino acid sequences of proteins occurs on aminoacyl-tRNA synthetases(for review, see[10]). Each of the 20 synthetases attaches its cognate amino acid to a specific tRNA producing AA-tRNAs. Elongation factor 1A.GTP (called EF-Tu;GTP in bacteria) binds the AA-tRNAs and transports them to the ribosomes for protein synthesis (reviewed in [11,12]). The critical fidelity of protein synthesis depends upon the remarkable ability of the synthetases to recognize their cognate amino acids and tRNAs and the ability of EF-Tu;GTP to discriminate between AA-tRNAs and tRNAs. The method of amino acid analysis proposed here is based on a simple idea: rather than separating the amino acids by chromatography, each synthetase or a tRNA specific for a different amino acid is separately positioned and the amino acids are analyzed from a mixture by monitoring the reactions catalyzed by the synthetases in a spatially resolved manner. Since EF-Tu;GTP specifically binds all AA-tRNAs but not tRNAs with high affinity, ligand assays can be developed to monitor the aminoacylation reactions catalyzed by the synthetases. By reducing amino acid analysis to the simplicity of a ligand assay and using spatially positioned biomolecular recognition molecules (synthetases, elongation factors, tRNAs) and high-specific-activity labels, massively paralleled amino acid analysis microarrays far more sensitive than current amino acid analyzers may be possible.

Aminopeptidases and carboxypeptidases are enzymes that release amino acids sequentially, one-at-a-time, from a protein's amino- or carboxy-termini, respectively. Over the years, these enzymes have been used in discontinuous kinetic assays for protein end-group sequencing. Due to the nonlinear rate of hydrolysis, these kinetic studies have been unsuccessful in most cases. I suggest a solution to this problem by creating real-time amino acid analysis microarrays that can detect amino acids as they are released by these enzymes. These arrays can be integrated on the same chip with aminopeptidase or carboxypeptidase digestions creating on-line microfluidic systems capable of generating either N- or C-terminal sequence data. As the amino acids are liberated by the exopeptidases, they flow form the site through the amino acid analysis microarrays and are detected sequentially. The terminal sequence is regenerated by a computer and the protein identified by automated database searching.

Principle of the method for amino acid analysis

All proteins from bacteria to man are synthesized from the same set of 20 amino acids. Each amino acid has one specific synthetase and one or more isoaccepting tRNA. The enzymes are highly selective for the amino acids they bind and even more selective in their attachment of the amino acids to their specific tRNAs. The basis of the method for amino acid analysis proposed here is that each synthetase and/or one isoaccepting tRNA specific for each amino acid is separated into a different chamber or immobilized onto a separate transducer or a spatially separated zone and the reactions catalyzed by the 20 synthetases are monitored using a position-sensitive detector. (fig. 1A). The extent of these reactions will be directly proportional to the concentration of the amino acids in solution. The amino acids (AAs) are being converted into AA-tRNAs by the following overall reaction:

(1.) AA + tRNA + ATP + Mg ++ <----------> AA-tRNA + AMP + PPi

There are a number of ways to follow this reaction. For example, one could include inorganic pyrophosphatase in the chambers and monitor the amount of inorganic phosphate produced in a continuous spectrophotometric assay (e.g. [13]). The reaction catalyzed by inorganic pyrophosphatase is

(2.) PPi + H_20 <---------->2Pi.

These assays can detect uM levels of phosphate and can be used to construct enzymatic amino acid analyzers. By using a macromolecule that specifically binds the AA-tRNAs, however, we can develop ligand assays to follow the aminoacylation reactions that are far more sensitive than these spectroscopic assays. EF-Tu;GTP binds all of the AA-tRNAs with high specificity forming a stable ternary complex (EF-Tu;GTP-AA-tRNA). So if EF-Tu;GTP is present with the synthetases and tRNAs the following reaction will occur :

(3.) EF-Tu;GTP + AA-tRNA <----------> EF-Tu;GTP-AA-tRNA

The ternary complex can be further stabilized by substituting a nonhydrolyzable GTP analog such as GDPNP [14,15]. Either EF-Tu;GTP or EF-Tu;GDPNP has a high affinity for all AA-tRNAs but effectively discriminates against deacylated tRNAs. The amount of ternary complex formed will be proportional to the concentration of the amino acid cognate to the tRNA and synthetase present. Hence, the amino acids will be identified and quantitated by following the formation of the ternary complex. This can be accomplished by labeling either the tRNAs, elongation factor, ternary complex, or a ternary complex probe and quantifying the label. Alternatively, the ligand assays can be based on label-free affinity biosensors (e.g. surface plasmon resonance-based biosensors). Ligand assays using fluorescent labels typically achieve femtomole-attomole sensitivities [16].

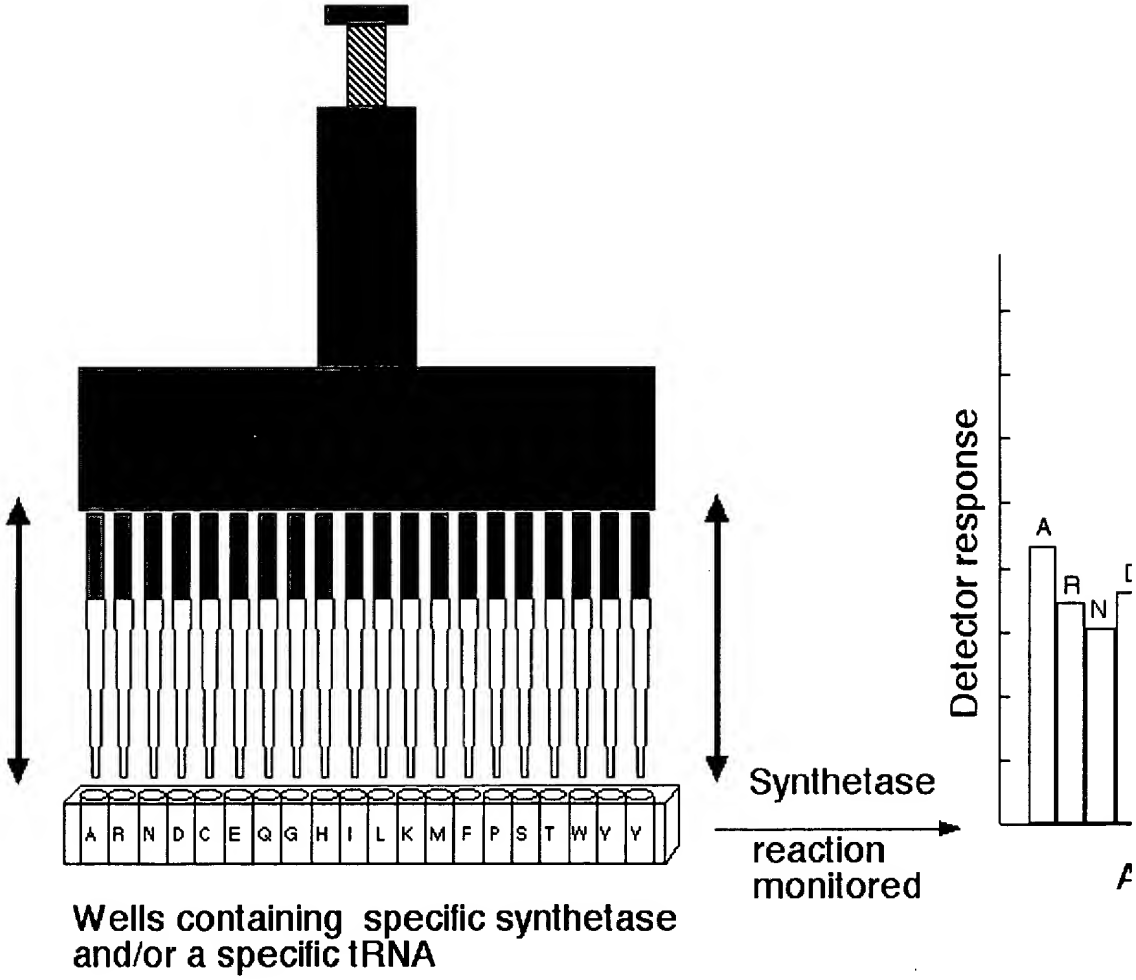
Key reactions

(1.) $AA + tRNA + ATP + Mg^{++} \rightleftharpoons$ AA-tRN

(2.) $PPi + H_2O \rightleftharpoons 2Pi.$

(3.) EF-Tu;GTP + AA-tRNA $\rightleftharpoons$ EF-Tu;G

A.



B.

Enzymatic amino acid analyzers . The key reactions for creating the enzymatic amino acid analyzers are shown above. The extent of these reactions will be proportional to the concentration of amino acids in solution.

A. Schematic illustration of a multichannel pipettor suitable for simultaneously mixing and dispensing a solution of the standard amino acids into 20 wells in a microtiter plate format. Each of the 20 aminoacyl-tRNA synthetases and/or a tRNA specific for a different amino acid is dispensed into a different well. The single letter codes for the amino acids are used here to represent the synthetase or tRNA cognate for a specific amino acid. Each well signals only the amino acid cognate to the synthetase and/or tRNA present. The reaction catalyzed by the synthetases is monitored in all wells simultaneously (using a plate reader) and), each synthetase will be dispensed into a separate well and the detector will be a spectrophotometer. used to quantitate the amino acids. If the synthetase reaction is followed using a continuous phosphate detection assay (reaction 2 and ref. 13)

B. An ultrasensitive ligand assay for the determination of the ternary complex in a microtiter plate format. In this scheme, EF-Tu;GTP is immobilized on the bottom of the wells. Each of the 20 wells in a row contains a mixture of the 20 synthetases in aminoacylation buffer and a fluorescently labeled tRNA specific for a different amino acid. After the amino acid mixture is added to the wells, the ternary complex will form on the bottom of the wells in response to the amino acids having a cognate tRNA in the well. After a washing step, the fluorescence is quantitated using a fluorometer (plate reader). Depending on the fluorescent labels and instrumentation these plate readers can achieve femtomole-attomole sensitivities (16).

Ultrasensitive ligand assays for amino acid analysis

The new idea here is to use separately positioned synthetases or tRNAs as a device for quantitatively discriminating among the 20 amino acids. The synthetase reactions can be followed in a number of ways (for example, see 13). Also, it is suggested to monitor the synthetase reaction by following the binding of AA-tRNAs to EF-Tu;GTP. It will be necessary to devise systems for following the formation of the AA-tRNA-EF-Tu;GTP complexes. Monitoring the binding of AA-tRNAs to EF-Tu;GTP using ligand assays is not new. For example, Janiak *et al* (17) used fluorescently labeled tRNAs to monitor the formation of the ternary complex in a continuous assay. They attached fluorescin covalently to the thiouridine at position 8, which is a conserved residue in all tRNAs. The fluorescently labeled tRNAs were functional both in aminoacylation by the synthetases and in binding to EF-Tu;GTP. A variety of ligand assay formats are possible for monitoring the binding of the AA-tRNAs to EF-Tu;GTP. By immobilizing one component of the ternary complex

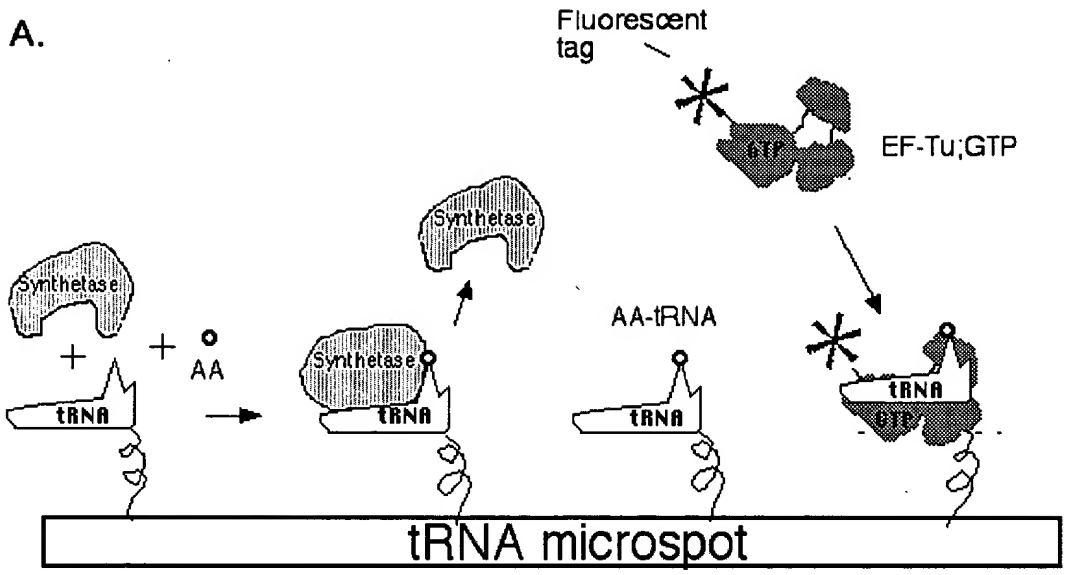
(either the tRNA or EF-Tu;GTP) and labeling the other component, the formation of the ternary complex can be followed as the labeled molecule is 'captured' onto the surface. The EF-Tu;GTP can be immobilized to a surface and the formation of the ternary complex on the surface can be monitored by capturing labeled tRNA molecules. (fig.1.B) Recently, immobilized EF-Tu;GTP has been used to purify AA-tRNAs by affinity chromatography(18,19). Like the free factor, the immobilized EF-Tu;GTP did bind AA-tRNAs, but not unacylated tRNAs. Hence, the immobilized elongation factor has been shown to retain its specificity for AA-tRNAs. Alternatively, the tRNAs (one specific for each amino acid) can be immobilized and the formation of the ternary complex on all 20 tRNAs can be monitored simultaneously. In this case, the ternary complexes can be detected with labeled elongation factors or labeled probes that bind the ternary complex. Other ligand assay formats can employ ternary complex probes (systems that specifically recognize charged AA-tRNA-EF-Tu;GTP complexes) to capture and/or detect the ternary complex. Macromolecules that specifically recognize the ternary complex include ribosomes and ribosomal subunits (e.g. cognate mRNA-charged small ribosomal subunits). It is also possible that antibodies or oligonucleotide aptamers that specifically bind the ternary complex can be produced.

Ultrasensitive ligand assays have relied on labeling one component of the reaction with a high-specificity-activity label (e.g. radioactive, fluorescent, or electrogenic labels) or a label that can participate in an amplification reaction. Label detection is the key determinant of sensitivity. Fluorescence is an outstandingly sensitive detection method, and laser-induced fluorescence has become the method of choice for single molecule detection (reviewed in 20). Since tRNAs,

elongation factors, ternary complexes or ternary complex probes can be labeled with multiple copies of fluorescent tags or with amplifying labels, it is possible to develop ligand assays for amino acid analysis with a sensitivity many orders of magnitude greater than exists for current methods of amino acid analysis.

Microarrays for amino acid analysis

Array technology is the method of choice for high throughput analysis. Miniaturized arrays are suggested for the quantitative analysis of many amino acid samples in parallel. Microarrays where thousands of different bioaffinity molecules are immobilized on a surface as spatially separated dots (usually 10-100um in diameter) and used to capture ligands have been developed for both nucleic acids and proteins (21,22). Array-based methods include sequencing by hybridization, cDNA expression profiling, comparative genome hybridization and genetic linkage analysis. Currently, the main large-scale application of microarrays is comparative gene expression analysis. Recently, whole genomes have been analyzed with cDNA or gene microarrays which capture fluorescently labeled mRNA (23). Microarray technology makes possible the development of amino acid analysis microchips. Two alternative routes are suggested to achieve the 20-fold spatial distribution of the recognition elements on microarrays. Either 20 specific tRNAs or the 20 specific synthetases are arrayed. Amino acid analysis can be performed with tRNA arrays by introducing a solution of the 20 synthetases in aminoacylation buffer along with the amino acid mixture to be analyzed. The AA-tRNAs formed on the array can be detected by binding a fluorescently labeled elongation factor. (fig. 2. A) Alternatively, each synthetase can be co-immobilized to a different microspot with EF-Tu;GTP to construct synthetase microarrays. (fig. 2B) Bound ternary complexes can then be detected by capturing fluorescently labeled tRNAs.



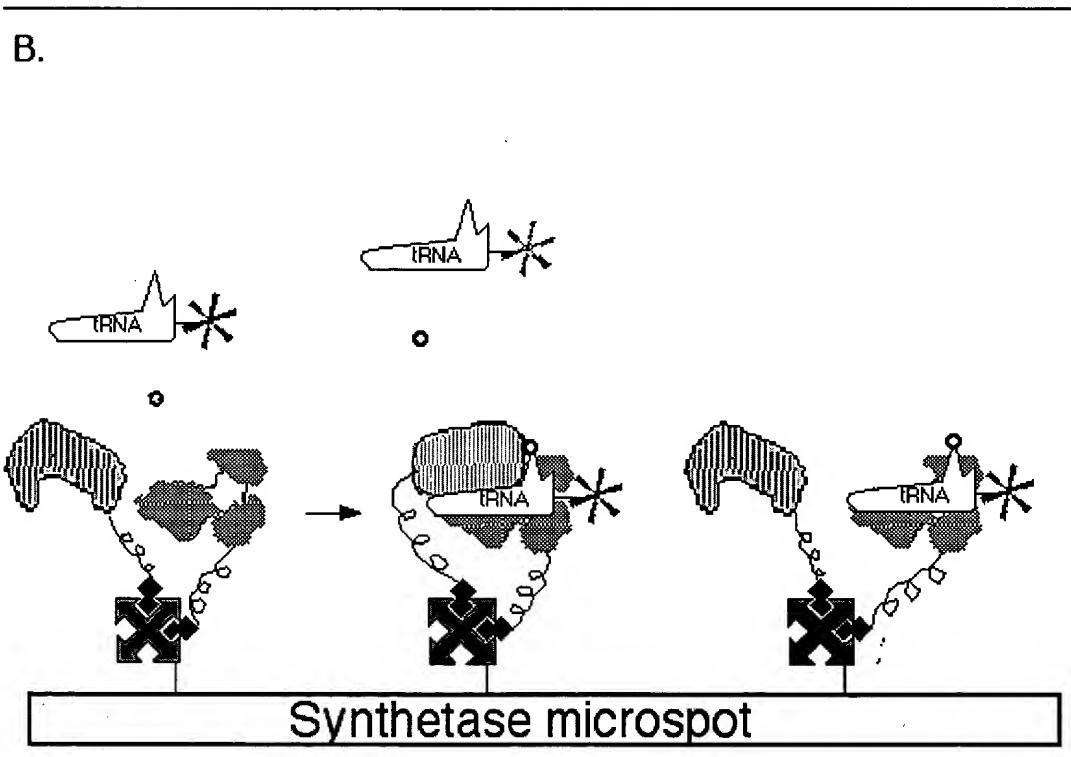


Figure 2

tRNA and synthetase microarrays.

A. tRNA microarray. Twenty tRNAs (one specific for each amino acid) are arrayed as microspots. A mixture of amino acids is added along with the 20

synthetases and fluorescently labeled EF-Tu;GTP in aminoacylation buffer. AA-tRNAs form on each microspot in proportion to the concentration of amino acid specific for the immobilized tRNA. Fluorescently labeled elongation factor binds the AA-tRNAs and is detected by laser induced fluorescence.

B. Synthetase microarrays. The 20 synthetases are co-immobilized with EF-Tu;GTP as separate microspots. Fluorescently labeled tRNAs are added along with the amino acid mixture. Amino acids and tRNAs specific for the microspot bearing their cognate synthetase will be converted into AA-tRNAs and 'captured' on the microspot by the EF-Tu;GTP. Amino acids will be detected by laser induced fluorescence.

Various methods for arraying both proteins and nucleic acids have been developed and will enable the arraying of the synthetases and tRNAs. For example, attachment of cDNA targets to a glass chip is usually carried out using non-covalent charge interactions. Glass slides have been coated with poly-L-lysine onto which clones are printed (24). This method of attachment relies on the electrostatic interactions between the positively charged lysine residues and the negatively charged phosphate backbone of the nucleic acids and hence should also be suitable to pattern the 20 specific tRNAs on a chip. Several methods of constructing protein microarrays have also been developed (25). Boehringer Mannheim developed a method that deposits arrays of 100-200 spots (~80 um in diameter) on the flat bottom of wells.

Ultrasensitive, spatially resolved detection

In addition to the requirement of high sensitivity, the enzymatic amino acid analyzers will require spatially resolved detection. Highly resolved optical detection of fluorescent labels enables the quantitation of captured ligands on arrays of many thousands of bioaffinity molecules simultaneously. Spatially resolved fluorescence detection in miniaturized systems can be achieved by direct imaging through a microscope, through optical fibers, or through optical waveguides. Fluorescent signals pass through spatial and spectral filters to the detection system. Ultrasensitive, spatially resolved detection on high-density arrays is typically achieved with confocal laser scanners or imaging CCD cameras (26,27). A schematic illustration of a possible amino acid analyzer microarray detection system is shown in fig. 3. Upon excitation by a laser, a CCD camera detects the pattern of fluorescent molecules on the surface, and image analysis software correlates the position of the fluorescent signal with the identity of the amino acids. Formation of the ternary complex on the surface can be measured quantitatively using a modified epifluorescence microscope equipped with a CCD camera. Real-time ultrasensitive measurements can be made, for example, by using evanescent wave excitation and an intensified video CCD camera. Many tiny spatially resolved wells or microflow channels can house the patterned array of recognition elements. Thousands of these chambers can be constructed on a single chip or plate. Fluorescent array detectors can quantitate end-labeled nucleotides at subattomole levels. A label-free probe can be detected using surface plasmon resonance (28), although array density and sensitivity are currently poor.

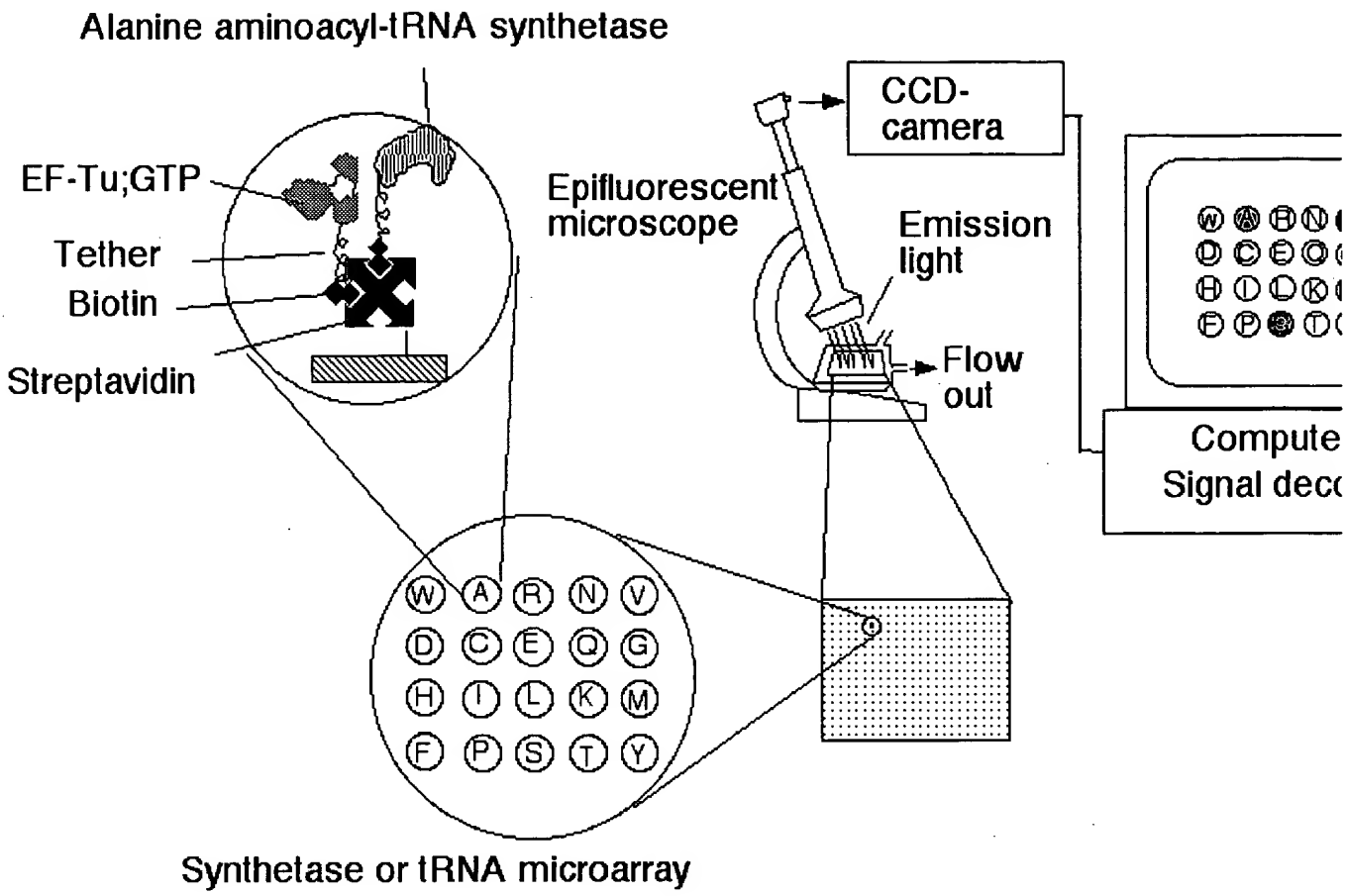


Figure 3

Spatially resolved detection system for amino acid analysis microarrays.

Patterned arrays of biorecognition elements (synthetases, or tRNAs) are immobilized on a surface at high spatial resolution as microdots. Single letter codes for the amino acids represent the synthetases or tRNAs specific for each amino acid. Each microdot will signal its cognate amino acid after it is captured into the ternary complex. The free member of the ternary complex (tRNAs or EF-Tu;GTP) is conjugated with fluorescent tags. Upon excitation of the fluorescent labels, a spatially specific detector allows analysis of the pattern of fluorescent molecules on the surface. As shown, the fluorescence is imaged through a microscope using a CCD detector. Fluorescence intensity on the arrays is correlated with the amount of amino acid in each sample. Thousands of samples may be analyzed on a single microchip.

A range of new micropipetting systems have been developed(29) and can be employed for delivery of microliter to nanoliter volumes of samples to the arrays. It is possible that carboxypeptidase or aminopeptidase digestons can be performed in nanowells and tiny samples aliquoted into amino acid analysis nanowells robotically during the time course of the digestion. These automated systems should be suitable for end-group sequencing of many samples in parallel at subpicomole levels. Due to the nonlinear rate of hydrolysis by these enzymes, continuous assays are preferable, however.

Towards an enzymatic nanosequenator

Carboxypeptidases and aminopeptidases have been used for sequencing proteins for many years(30-32). During the course of a digestion, samples are taken and analyzed later using an amino acid analyzer. Inherent in the use of these discontinuous assays is the assumption that the amino acids are being released linearly with time over the period chosen, however, the rate of cleavage is sequence dependent and varies unpredictably. When an analyte varies unpredictably a continuous (real-time) assay is needed. A continuous assay is not possible

using existing methods of amino acid analysis. On the other hand, continuous amino acid biosensor arrays can be developed using the method of amino acid analysis described here. Indeed, harnessing the power of affinity reactions for real-time and on-line monitoring has been one of the major driving forces for biosensor research.

Amino acid analysis biosensor arrays

The formation of macromolecular complexes, for example the formation of the ternary complex, can now be studied in real-time using miniaturized biosensors (reviewed in, [33-34]). A biosensor is a self-contained integrated device that is capable of providing quantitative or semi-quantitative analytical information using a biological recognition element which is in direct contact with a transduction element. Biosensors can be classified according to their transduction mechanisms and include microelectrodes, surface acoustic wave sensors, and fiber optic sensors. A commercially available biosensor system called BIAcore (Pharmacia Biosensor, Uppsala, Sweeden) contains a sensor microchip, a laser light source emitting polarized light, an automated fluid handling system, and a diode-array position sensitive detector[35]. This system uses a surface plasmon resonance-based assay, an optical technique that measures changes in the refractive index at the sensor chip surface. These systems can monitor biological interaction phenomena at surfaces in real-time under continuous flow conditions. Any of the usual energy transduction modes can be fabricated in an array format and used to construct amino acid analysis biosensor arrays. Each biorecognition element can be placed on transducers which monitor mass changes, the formation of electrochemical products, or the presence of fluorescence. Optical and electrochemical transducers, however, provide the most sensitive biosensors and are well suited for miniaturization.

Fiber optics can be used as thin flexible pipes to transport light to and from tiny volumes of immobilized chemistry at the probe end. Optical fibers offer several advantages for the construction of amino acid analysis biosensor arrays. Walt's group recently pioneered the use of optical fiber bundles for combining sensing and imaging and creating optical sensing arrays [36]. Optical fibers provide a highly miniaturizable transduction format and may allow monitoring in real-time [37]. The tips of optical fibers can be of submicron dimensions. Optical fiber sensors can use the evanescent field to excite and collect the fluorescence of molecules bound to the surface. The evanescent wave excites only fluorophores that are bound to the surface. This allows real-time detection of the captured probe on microarrays even in the presence of high concentrations of fluorophores in the bulk solution without washing. Specific synthetases or tRNAs cognate for a different amino acid can be immobilized to a separate fiber. Each tiny fiber tip will signal the amino acid cognate to the synthetase or tRNA immobilized. For example, each synthetase can be co-immobilized with EF-Tu;GTP to a different fiber, and the formation of the ternary complex on each fiber can be monitored using fluorescently labeled tRNAs . When coupled to aminopeptidase or carboxypeptidase digestions these ultrasmall sensors may allow real-time protein end-group sequencing in tiny volumes (<1ul). It will be necessary to design the system so that each liberated amino acid will diffuse (or be transported) to its immobilized synthetase (or tRNA) and be detected before the next amino acid is released. Diffusion is rapid over distances of a few micrometers. Diffusion times increase with the size of the molecule and the square of the distance traveled. For amino acids in water at 37o C diffusion times are approximately 1msec for 1um, 0.1 sec for 10um and 17min for 1 mm. For systems larger than a few micrometers, transport must take place by convection for real-time sequential monitoring. A possible C-terminal enzymatic protein sequenator with an evanescent wave optical fiber amino acid analyzer biosensor array is illustrated in fig.4. As shown the carboxypeptidases are immobilized to a separate zone within

the nanochamber. The carboxypeptidase digestion chamber can be separated from the biosensor array by a semipermeable membrane having a molecular weight cutoff that allows amino acids to pass freely but is impermeable to macromolecules. As amino acids are released into solution by the carboxypeptidases they cross the membrane and are detected by the biosensor array. Amino acid analyzer arrays can be incorporated into microdialysis or ultrafiltration probes and coupled to amino- or carboxypeptidase digestions for end-group sequencing. Indeed, flow systems offer many advantages for analysis.

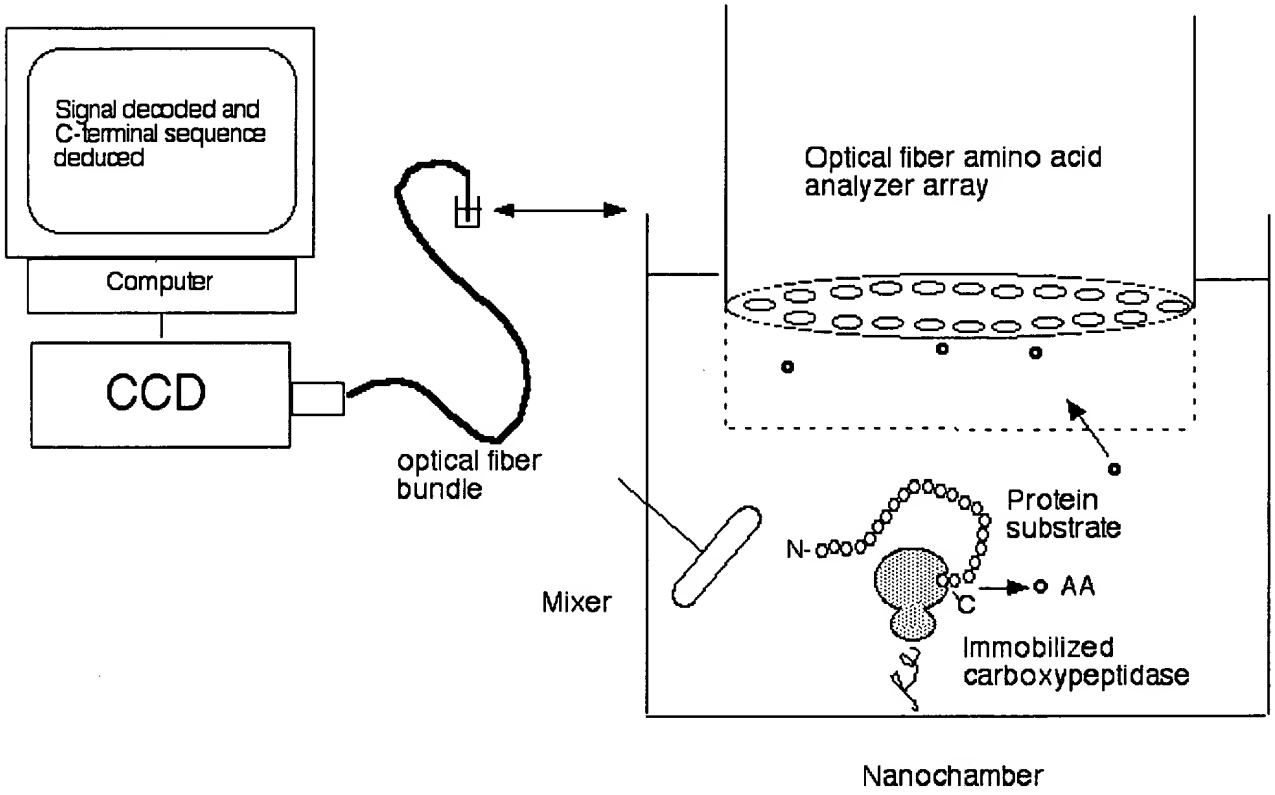


Figure 4

Schematic diagram an evanescent-wave optical fiber C-terminal nanosequenator. Each synthetase is co-immobilized with EF-Tu;GTP to a different optical fiber. The tRNAs are labeled with fluorescent tags, and the formation of the ternary complex on the sensor surface is monitored continuously by fluorescence detection. The biorecognition elements (synthetases, EF-Tu;GTP, tRNAs) are enclosed by a semipermeable membrane having a molecular weight cutoff such that it allows free passage of amino acids but is impermeable to macromolecules. These miniaturized optical fiber amino acid biosensor arrays may be coupled with aminopeptidases or carboxypeptidases for protein end-group sequencing. Fluorescently labeled tRNAs will be captured on the sensors surface in response to amino acids cognate to the immobilized synthetases and the increase in fluorescence on each fiber will be plotted as a function of time to regenerate the terminal sequence of the protein.

Spatially resolved microflow systems

Because diffusion in liquids is slow (over distances greater than a few micrometers) and aimless, the incorporation of arrays into flow systems for automated processing will facilitate high throughput analysis and permit sequential monitoring. Solid-phase ligand assays are currently performed in microtiter plates; however, this technique requires long incubation times to achieve equilibrium conditions and is difficult to miniaturize and automate. On the other hand, flow systems are easily automated and miniaturized and allow fine control of reagent additions and rapid chemistries by reducing diffusional limitations. In addition, reproducibility is extremely high and calibrations are easy to perform(38). When coupled with microdialysis and

flow injection systems, biosensors have become available for on-line, real-time monitoring [39-43]. It should be possible to couple amino acid biosensor arrays to microdialysis or miniaturized ultrafiltration systems to achieve real-time and ultrasensitive detection of amino acids as they are released into solution by carboxypeptidases or aminopeptidases.

Progress in the computer industry has made possible micromachining of silicon into complex shapes with dimensions of less than 1um creating an ever increasing number or miniaturized devices. Devices can be electrical, such as microelectrodes, and signal transducers; optical such as photodiodes and optical waveguides; and mechanical, such as pumps and microreactors. In the emerging field of microfluidics, the integration of automated microflow devices and sensors allow very precise control of ultrasmall flows on microchip platforms[44,45]. Micrometer-scale pumps and valves transport liquid solutions from separate reservoirs through channels of micrometer dimensions. Many different flows can be combined in all sorts of ways and mixed on the same chip. Chip technology also allows the integration of intersecting channels, reaction chambers, mixers, filters, heaters, and detection devices to perform on-chip reactions in sub-nanoliter volumes in a highly controlled and automated fashion[46,47]. Microflow channels have very low dead volumes allowing rapid chemistry and detection and real-time monitoring. In addition, if reagents are continuously replenished by micropumps, sequential measurements are possible. By microfabrication of many parallel channels on a single microchip platform, multiple samples can be analyzed simultaneously. Laser-induced fluorescence is generally the detection method of choice for microflow devices. Ultrasensitive spatially specific detection in capillary or microchannel arrays can be achieved with laser-induced fluorescence using imaging detectors or confocal scanners (reviewed in, [48]). Integrated, on-line microfluidic systems can be developed to transport the amino acids and other reagents from separate chambers through the amino acid analysis microarrays allowing sequential measurements and rapid (possibly direct) data acquisition.

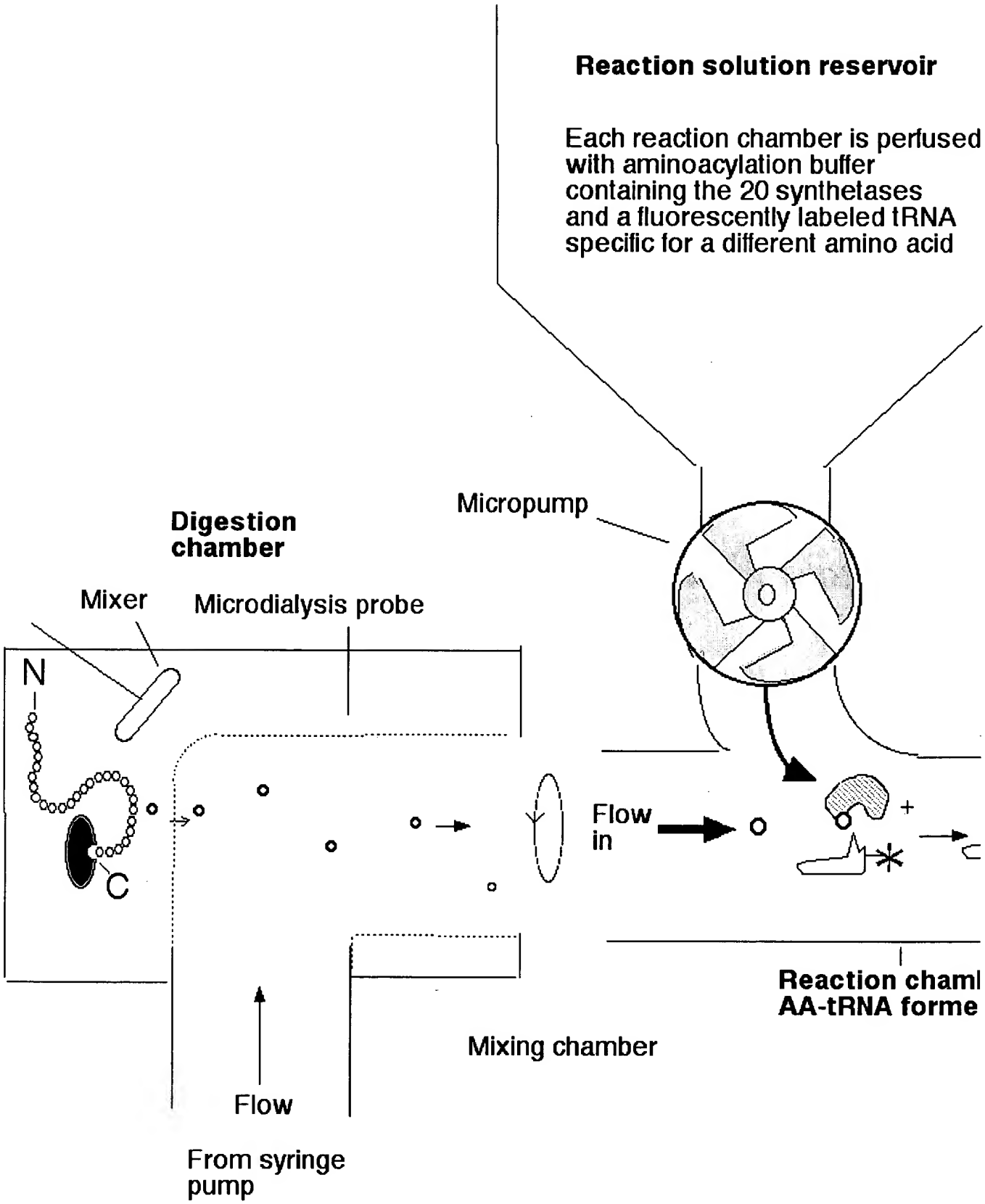
Continuous-flow protein sequencing on a chip

Recent developments in microfabricated devices will make possible the construction of integrated microchip enzymatic nanosequenators. The relevant components such as continuous flow mixers, pumps, microreactors, microdialysis and miniaturized ultrafiltration systems have already been developed [49]. A possible C-terminal nanosequenator is illustrated in fig. 5A. The carboxypeptidase digestion takes place in a tiny chamber traversed by a microdialysis probe or miniaturized ultrafiltration probe. The membrane has a molecular weight cutoff such that it allows free passage of amino acids but is impermeable to the protein substrate and carboxypeptidases. As the amino acids are liberated from the protein's termini, they cross the membrane and enter a flow stream which carries them to the biosensor array for continuous detection. The record of the continuous detection of the 20 amino acids by the biosensor array will give the terminal sequence of the protein.

As illustrated in fig. 5A, the biosensor array consists of 20 reaction channels. Each reaction channel has an inlet for amino acids to flow into and an outlet and is connected to a different reservoir by a microchannel. Each reservoir contains the 20 synthetases and a fluorescently labeled tRNA specific for a different amino acid. (Alternatively, each reservoir can contain one synthetase and a fluorescently labeled cognate tRNA). The 20 reaction solutions are continuously transported through the reaction channels (e.g. by continuous-flow micropumps). EF-Tu;GTP is immobilized downstream in each reaction channel and used to capture and detect the fluorescently labeled AA-tRNAs. A laser beam is focused on the immobilized EF-Tu;GTP. For example, EF-Tu;GTP can be immobilized to optical fibers placed in each reaction

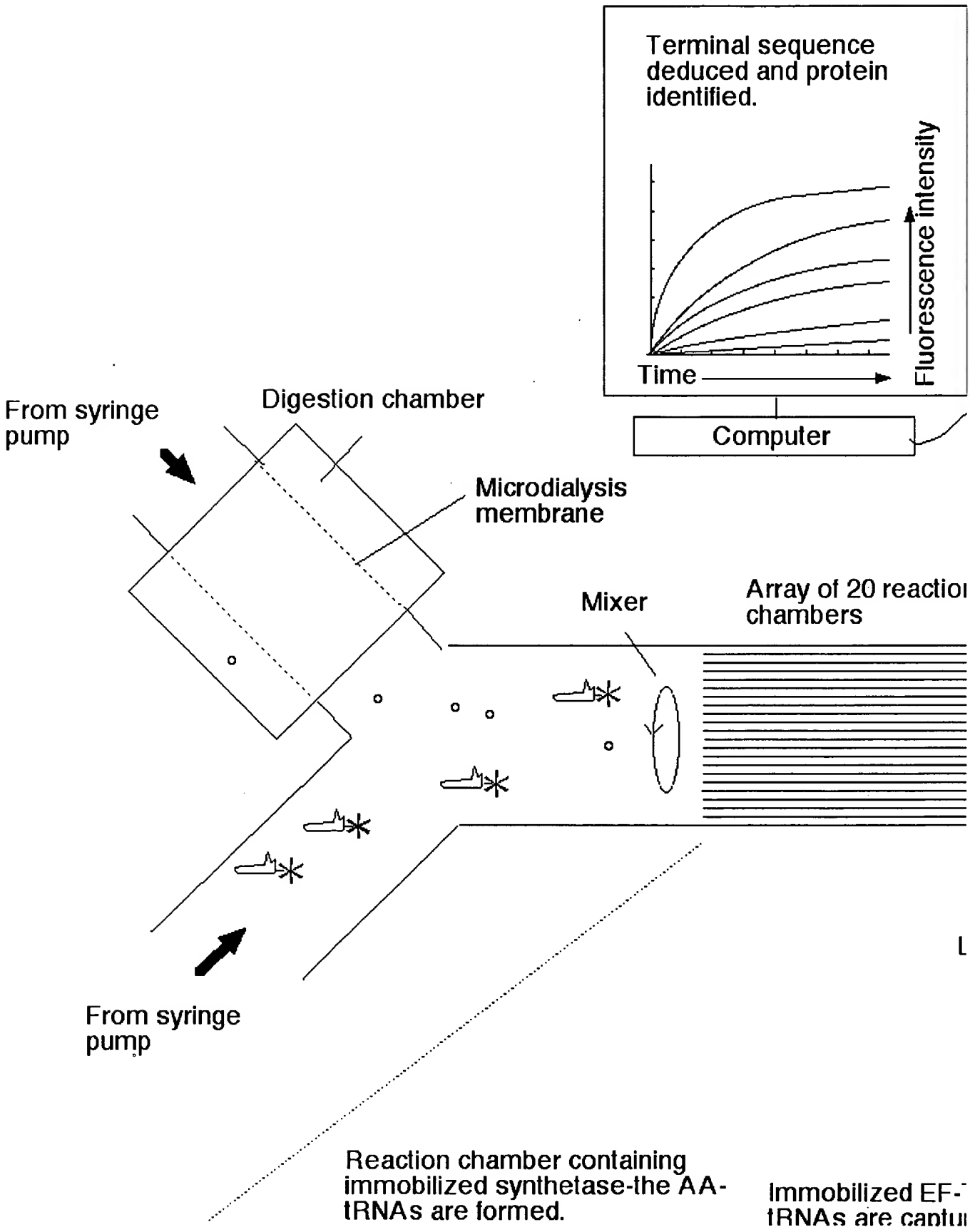
channel. It is possible to selectively excite and collect fluorescence from the immobilized AA-tRNA-EF-Tu;GTP complexes even in the presence of high concentrations of fluorescently labeled tRNAs in the bulk solution (e.g. using evanescent wave excitation). As amino acids cognate for the fluorescently labeled tRNA in each channel flow through, a proportionate amount of fluorescently labeled AA-tRNA will be formed. The newly formed AA-tRNAs will be captured downstream by the immobilized elongation factor. When AA-tRNAs bind to the immobilized elongation factor laser light in the evanescent wave will excite the fluorophore generating a signal. The GTP bound elongation factor binds AA-tRNAs but the GDP bound elongation factor does not. It should be possible to elute the AA-tRNAs with a buffer containing GDP and to reactivate the immobilized elongation factor-based sensor with a regeneration buffer containing GTP. This method has been used to affinity purify AA-tRNAs on immobilized EF-Tu;GTP[50].

Other microflow formats are possible. For example, either a different synthetase (fig. 5B) or tRNA (one specific for a different amino acid) can be immobilized in the 20 reaction channels. Alternatively, an amino acid biosensor array (e.g. optical fiber or microelectrode) can be placed in a flow stream and detect amino acids as they flow through the array. (fig. 5C).


Reaction solution reservoir
Each reaction chamber is perfused with aminoacylation buffer containing the 20 synthetases and a fluorescently labeled tRNA specific for a different amino acid
Micropump
Digestion chamber
Mixer
Microdialysis probe
N
C
Flow in
+
Reaction chaml
AA-tRNA forme
Mixing chamber
Flow
From syringe pump

Continuous-flow protein sequencing on a chip.

A. Schematic drawing of a continuous flow C-terminal nanosequenator. The carboxypeptidase digestion takes place in a tiny chamber and free amino acids are collected by a microdialysis or ultrafiltration probe. As the amino acids are released by the carboxypeptidase, they cross the membrane and enter the flow stream. This flow stream enters a continuous flow mixer and then an array of 20 reaction channels. Only one reaction chamber is shown. In each of the 20 channels, the flow stream is joined by a second flow stream carrying fluorescently labeled tRNAs specific for a different amino acid and a mixture of the 20 synthetases. EF-Tu;GTP is immobilized downstream in each reaction channel. Amino acids cognate for the labeled tRNA in each channel will be converted into a proportionate amount of AA-tRNA which will be captured by the immobilized EF-Tu;GTP. A laser beam focused on the immobilized EF-Tu;GTP will allow detection of the fluorescently labeled AA-tRNAs as they bind the elongation factor. The record of the binding of AA-tRNAs to EF-Tu;GTP in the 20 chambers will give the C-terminal sequence of the protein.


Terminal sequence deduced and protein identified.
Fluorescence intensity
Time
Computer
From syringe pump
Digestion chamber
Microdialysis membrane
Mixer
Array of 20 reactio chambers
From syringe pump
Reaction chamber containing immobilized synthetase-the AA-tRNAs are formed.
Immobilized EF-
tRNAs are captu

B. In this format, each synthetase is immobilized in a separate reaction channel. Amino acids sequentially liberated from the protein's end traverse the microdialysis probe and enter the flow stream which is joined by a second flow stream carrying fluorescently labeled tRNAs (one cognate to each amino acid) in reaction buffer. The two flows are mixed on the chip and enter the reaction channel array. In each reaction channel, amino acids having a cognate synthetase immobilized will be converted into a proportionate amount of AA-tRNA. The newly formed AA-tRNA will bind to EF-Tu;GTP immobilized downstream and be continuously detected as described above.

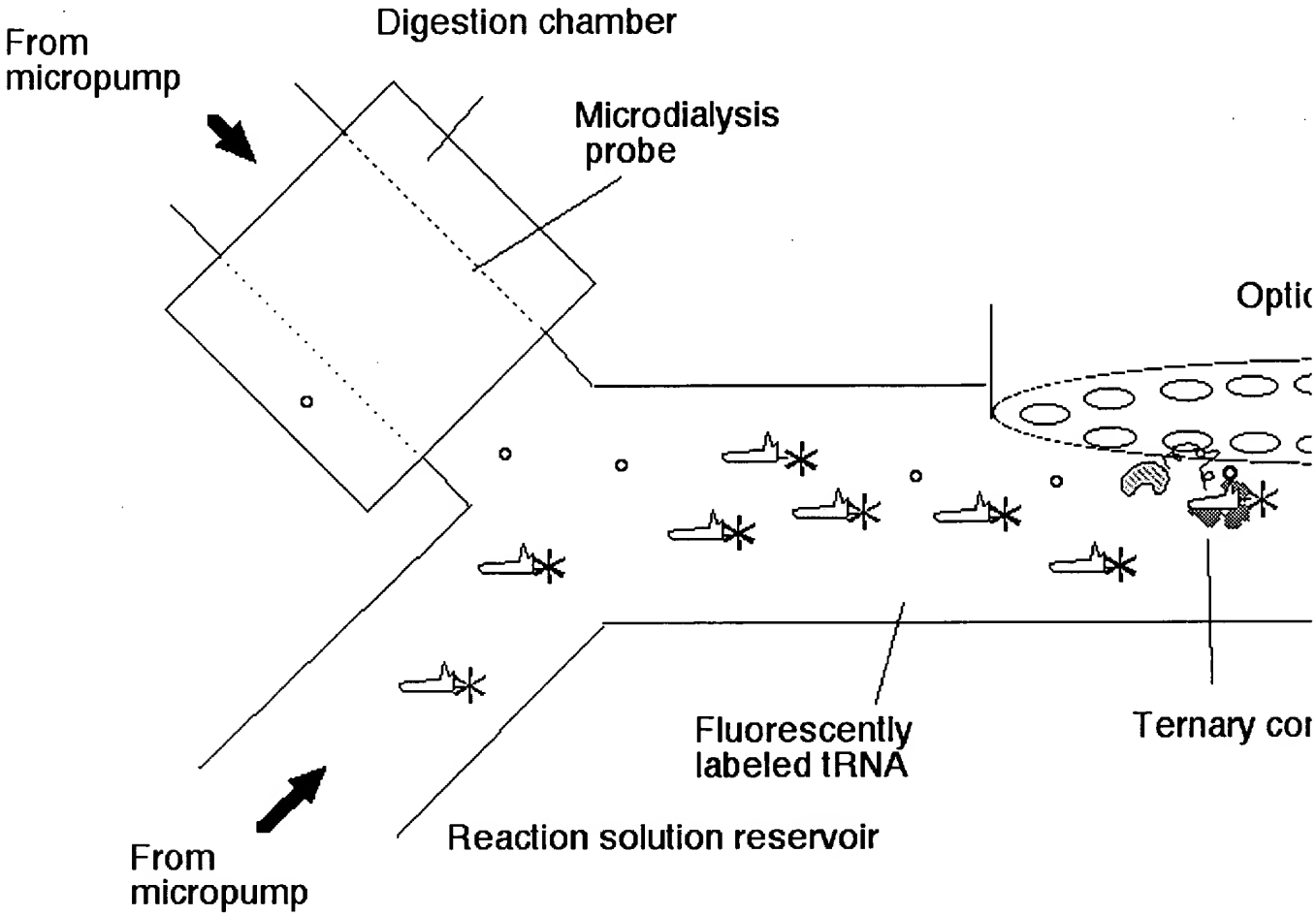


Figure 5c

Schematic drawing of a flow optical fiber nanosequenator.

Two flow streams are combined and mixed on the chip. One stream contains exopeptidase digestion buffer and is pumped through a microdialysis probe that traverses a tiny digestion chamber. Liberated amino acids cross the microdialysis membrane and enter the flow stream. The second flow stream contains a mixture of fluorescently labeled tRNAs (one specific for each amino acid) in aminoacylation buffer. The two flows are combined and mixed on the chip. This stream containing the mixture of amino acids and fluorescently labeled tRNAs flows through the biosensor array. Amino acids are attached to their cognate tRNAs by the immobilized synthetases and bind co-immobilized elongation factors. The bound AA-tRNAs are then quantitated by fluorescence detection.

Prospects for proteomics

With the ability to purify thousands of proteins in a single experiment by 2D gel electrophoresis, there is an urgent need for rapid methods of protein identification and analysis. There has been much interest lately in using amino acid analysis for rapid and inexpensive protein identification (51-53). In analogy to gene expression profiling on cDNA microarrays, high-throughput amino acid analysis microchips have been described here. These microarrays should allow amino acid analysis of all proteins separated by a 2D

electrophoresis gel on a single chip or plate simultaneously. Protein sequencing is pivotal to biological research. It is especially important to determine the N- and C-terminal sequence of an intact protein. End-group sequencing will identify the start and stop point of a protein or gene; may allow the PCR cloning of the intact gene; will identify limited proteolysis, a common and important regulatory mechanism; and will provide a powerful method to identify proteins separated by 2D gels. Microfabricated end-group sequenators such as those described can be constructed as massively paralleled computer controlled and integrated systems where both N- and C-terminal sequencing of many proteins can be performed on a single chip platform simultaneously. The terminal sequence tags generated can be processed on-line and the proteins identified by data base searching. Among the most outstanding deficiencies in the current set of methods in protein chemistry are the ones for C-terminal sequencing. Since no sensitive, reliable method for C-terminal sequencing is available, the C-terminus of proteins is a region that is often not analyzed. Hence, the C-terminal sequenators proposed here will be especially useful in proteome projects.

Another important consideration in proteome studies is sensitivity. Proteins, unlike genes, have no amplification methods such as PCR or cloning, so sensitivity is even more important than for gene analysis. In order to analyze minor proteins separated from 2D gels we must work at the femtomole level. Current amino acid analyzers and sequenators fall short of this level by three orders of magnitude. The Nobel Prize-winning amino acid analyzers of Stanford Moore and William Stein, where amino acids are separated by automated chromatography and quantitated by reacting them with ninhydrin, first described more than forty years ago, are still the most common method of amino acid analysis. These instruments only achieve nanomole sensitivity. By converting the amino acids into macromolecules (AA-tRNAs) that are specifically bound by another macromolecule (EF-Tu;GTP) with a high affinity, it will possible to devise ultrasensitive ligand assays and automated microarrays for amino acid analysis. With laser induced fluorescence as a detection method, it should be possible to create simple amino acid analyzers and end-group sequenators with a sensitivity approaching the single molecule level.

The development of these miniaturized array-based systems should improve the throughput and sensitivity of amino acid analysis and protein end- group sequencing by orders of magnitude and allow proteins, like DNA and RNA, to be subjected to mass screening.

I would like to thank Joshua Deutsch, Eliot Dresselhaus, Jon Harada, and Russell Doolittle for help with the manuscript. Thanks to Yana Ning for drawing the figures and to Leroy Hood and Sydney Brenner for reading an earlier version of the manuscript.

This work is subject to a patent in process.

References

1. Watson, A., Mazumder, A., Stewart, M. , and Balasubramanian, S. (1998). Technology for microarray analysis of gene expression. Curr. Opion. Biotech. **9**, 609-614.

2. Lipshutz, R. J., Fodor, S. P. A., Gingeras, T. R., and Lockhart, D. J

(1999). High density synthetic oligonucleotide arrays. Nat. Genet. **21**, 20-24.

3. McKenzie, S. E., Mansfield, E., Rappaport, E., Surrey, S., and Fortina, P. (1998). Parallel molecular genetic analysis. Eur. J. Hum. Genet. **6**, 417-429.

4. Zhang, N. Tan, H., and Yeung, E.S. (1999). Automated and integrated system for high-throughput DNA genotyping directly form blood. Anal. Chem. **71**, 1138-1145.

5. Wooley, A. T. et al (1996). Functional integration of PCR amplification and capillary electrophoresis on a microfabricated DNA analysis device. Anal. Chem. **68**, 720-723.

6. Wilkins, M. R. , Williams, K. L., Appel, R.D. and Hochstrasser, D. F. (ed.) (1997). *Proteome research: new frontiers in functional genomics.* New York: Springer.

7. O'Farrell, P. H. (1975). High resolution two-dimensional electrophoresis of proteins. J. Biol. Chem. **250**, 4007-4021.

8. Wilkins, M. R., Pasquali C., Appel, R.D., Ou, K., Golaz, O., Sanchez, J., C., Yan, J. X, Gooley, A. A, Hughes, G., Humphery-Smith, I., Williams, K. L, and Hochstrasser, D. F. (1996). From proteins to proteomes: large scale protein identification by two-dimensional electrophoresis and amino acid analysis. Biotechnology. **14**, 61-65.

9. Wilkins, M. R. et al (1998). Protein identification with N and C-terminal sequencing tags in proteome projects. J. Mol. Biol. **278**, 599-608.

10. Schimmel, P. (1987). Aminoacyl tRNA synthetases: General scheme of structure-function relationships in the polypeptides and recognition of transfer RNAs. Ann. Rev. Biochem. **56**, 125-158.

11. Nissen, P., Kjeldgaard, M. Thriup, S., Clark, B. F. C., and Nyborg, J. (1996). The ternary complex of aminoacylated tRNA and EF-Tu;GTP: Recognition of a bond and a fold. Biochime, **78**, 921-933.

12. Clark, B. (1997). The ternary complex of EF-Tu and its role in protein biosynthesis. Curr. Opin. Struct. Biol. **1**, 110-116.

13. Lloyd, A. J., Thomann, H. U., Ibba, M., and Soll, D. (1995). A broadly applicable continuous spectrophotometric assay for measuring aminoacyl-tRNA synthetase activity. Nucleic Acids Res. **23**, 2886-2892.

14. Wagner, A., Simon, I., Sprinzl, M., and Doody, R. S. (1995). Interaction of Guanosine nucleotides and their analogs with elongation factor Tu from *Thermus thermophilus.* Biochemistry. **34**, 12535-12542.

15. Nissen, P. et al (1995). Crystal Structure of the ternary complex of Phe-tRNAphe, EF-Tu, and a GTP analog. Science. **270**, 1464-1472.

16. Zubritsky, E., (1999). Microplate fluorometers reach critical mass. Anal. Chem. **71**, 39A-43A.

17. Janiak F., Dell, V.A., Abrahamson, J.,K., Watson, B.,S., Miller, D. L., and Johnson, A.,E. (1990). Fluorescence characterization of the interaction of various transfer RNA species with elongation factor Tu.GTP: evidence for a new functional role for elongation factor Tu in protein biosynthesis. Biochemistry **29**, 4268-4277.

18. Robeiro, S., Nock, S., and Sprinzl, M. (1995). Purification of aminoacyl-tRNA by affinity chromatography on immobilized *Thermus thermophilus* EF-Tu.GTP. Anal. Biochem. **228**, 330-335.

19. Chinali, G. (1997). Isolation of tRNA isoacceptors by affinity chromatography with immobilized elongation factor Tu from *Escherichia coli.* J. Biochem. Biophys Methods. **34**, 1-10.

20. Nie, S. and Zare, R.N. (1997). Optical detection of single molecules. Ann. Rev. Biophys. Biomol. Struct. **26**, 567-96.

21. Fodor, S. P. A., Read, J. L., Pirrung, M. C., Stryer, L., Lu, A. T., and Solas, D. (1991). Light-directed spatially directed paralleled chemical synthesis. Science **251**, 767-773.

22. Lueking, A., Horn, M., Eickhoff, H., Bussow, K, Lehrach, H., and Walter, G. (1999) Protein microarrays for

gene expression and antibody screening. Anal. Biochem. **270**, 103-11.

23. Wodicka, L., Dong, H., Mittmann, M., Ho, M., H., and Lockhart, D. J. (1997). Genome wide expression monitoring in Saccharomyces cerevisia. Nat. Biotechnol. **15**, 1359-1367.

24. DeRisi, J. L. Lyer, V. L. Brown, P. O. (1997). Exploring the metabolic and genetic control of gene expression on a genomic scale. Science. **278**, 680-686.

25. Ekins, R. P. (1998). Ligand assays: from electrophoresis to miniaturized microarrays. Clin. Chem. **44**, 2015-2030.

26. Guschin, D., et al. (1997). Manual manufacturing of oligonucleotide, DNA, and protein microchips. Anal Biochem, **250**, 203-211.

27. Bowtell, D. D. L., (1999). Options available-from start to finish-for obtaining expression data by microarray. Nat. Genet. **21**, 25-32.

28. Thiel, A.J., Frutos, A. G., Jordan, C. E., Corn, R. M., and Smith, L. M., (1997). *In situ* surface plasmon resonance imaging detection of DNA hybridization to oligonucleotide arrays on gold surfaces. Anal. Chem. **69**, 4948-4956.

29. Rose, D., and Lemmo, A.V. (1997). Challenges in implementing high-density formats for high throughput screening. Automat. News. **2**, 12-19.

30. Light, A. (1967) Leucine Aminopeptidase. Methods Enzymol. **11**, 426-444.

31. Breddam, K., and Ottesen, M. (1987). Determination of C-terminal sequences by digestion with serine carboxypeptidases: the influence of enzyme specificity. Carlsberg Res. Commun. **52**, 55-63.

32. Royer, G. P. (1973). Immobilized derivatives of leucine aminopeptidase and aminopeptidase M. J. Biol. Chem. **218**, 1807-1812.

33. Rogers, K. R., and Mulchandani, A. (eds) (1998). *Affinity biosensors: techniques and protocols.* Totowa, New Jersey: Humnn Press.

34. Cunningham, A. J. (1998). *Introduction to bioanalytical sensors.* John Wiley and Sons. N. Y.

35. Raghavan, M., and Bjorkman, P. L. (1995). BIAcore: a microchip-based system for analyzing the formation of macromolecular complexes. Structure. **3**, 331-333.

36. Healey, B. G., Foran, S. E., and Walt, D. R. (1995). Photodeposition of micrometer-scale polymer patterns on optical imaging fibers. Science. **269**, 1178-1180.

37. Healey, B. G., and Walt, D. R. (1997). Fast temporal response fiber-optic chemical sensors based on the photodeposition of micrometer-scale polymer arrays. Anal. Chem. **69**, 2213-2216.

38. Scheller, F.,W., Schnbert, F., and Fedrowitz, J. (eds.) (1997). *Frontiers in Biosensors 1 Fundamental Aspects.* Basel Switzerland: Birkhausen Verlan.

39. Freaney R., et al. (1997). Novel instrumentation for real-time monitoring using miniaturized flow systems with integrated biosensors. Ann. Clin. Biochem. **34**, 291-302.

40. Cook, C. J. (1997). Real-time measurements of corticosteroids in conscious animals using an antibody-based electrode. Nat. Biotech. **15**, 467-71.

41. Steele, K. M., and Lunte, M. E. (1995). Microdialysis sampling coupled to on-line microbore liquid chromatography of pharmacokinetic studies. J. Pharm. Biomed Anal. **13**, 149-154.

42. Kaptein, W. A., et al. (1997). Methodological aspects of glucose monitoring with a slow continuous subcutaneous and intravenous ultrafiltration system in rats. Biosens. Bioelectron. **12**, 967-976.

43. Niwa, O., Torimitsu, K., Morita, M., Osborne, P. and Yamamoto, K. (1996). Concentration of extracellular L-glutamate released from cultured nerve cells measured with a small volume on-line sensor. Anal. Chem. **68**, 1865-1870.

44. Gravesen, P., Branebjerg, J.. and Jensen, O. S. (1993). Microfluidics-a review. J. Micromech. Microeng. **3**, 168-182.

45. Shoji, S. and Esashi, M. (1994). Microflow devices and systems. J. Micromech. Microeng. **4**, 157-171.

46. Colyer, C., L., Tang, T., Chen, N., and Harrison, D. J. (1997). Clinical potential of microchip capillary electrophoresis systems. Electrophoresis. **18**, 1733-1741.

47. Effenhauser, C. S., Bruin, G. J. M., and Paulus, A. (1997). Integrating chip-based capillary electrophoresis. Electrophoresis. **18**, 2203-2213.

48. Kheterpal, I. and Mathise, R. A. (1991). Capillary array electrophoresis DNA Sequencing. Anal. Chem. **71**, 31A-37A.

49. Kricka, L. J., (1998). Miniaturization of analytical systems. Clin. Chem. **44**, 2008-2014.

50. Chinali, G. (1997). Isolation of tRNA isoacceptors by affinity chromatography with immobilized elongation factor Tu from *Escherichia coli*. J. Biochem. Biophys Mehtods. **34**, 1-10.

51. Yan, J. X., Wilkins, M. R., Ou K, Gooley A.A, Williams, K. L., Sanchez, J. C. Gloaz, O., Pasquali, C., and Hochstrasser, D. F. (1996). Large-scale amino-acid analysis for proteome studies. J. Chromatogr A. **736**, 291-302.

52. Wilkins, M. R, and Williams, K. L. (1997) Cross-species protein identification using amino acid composition, peptide mass fingerprinting, isoelectric point and molecular mass: a theoretical evaluation. J. Theor. Biol. **186**,7-15.

53. Shaw, G. (1993) Rapid identification of proteins. Proc. Natl. Acad. Sci. USA. **90**, 5138-5142.